Comparative Effectiveness Trial Between a Clinic- and Home-Based Complementary and Alternative Medicine Telerehabilitation Intervention for Adults with Multiple Sclerosis (MS)

NCT #: NCT03117881

TEAMS Study Protocol and Statistical Analysis

Plan Document Version Date: August 20th, 2022

METHODS

Study Overview

The TEAMS study was a cluster randomized, comparative effectiveness trial comparing the effectiveness of TeleCAM and DirectCAM delivered at 43 clinics across Alabama, Mississippi, and Tennessee. There were 40 clinics initially randomized and 3 satellite sites including Cookeville, Ashland City and Mobile were added later. These 3 satellite sites were not randomized as they were close to and were associated with support groups at the existing randomized sites. The satellites sites were recruited to alleviate logistical challenges in throughput around baseline testing after enrollment.

Using the clinics as the units of randomization, our criteria while identifying clinic sites were to minimize the risk of contamination while having enough sites for a cluster randomization, ability to detect a small to moderate effect size, and ability to recruit efficiently. The randomization list was run by a study statistician who was not involved in any participant interaction and intervention delivery. Given the nature of this study design, group assignments were not blinded to participants and study staff involved in the intervention delivery but were blinded to data analysts.

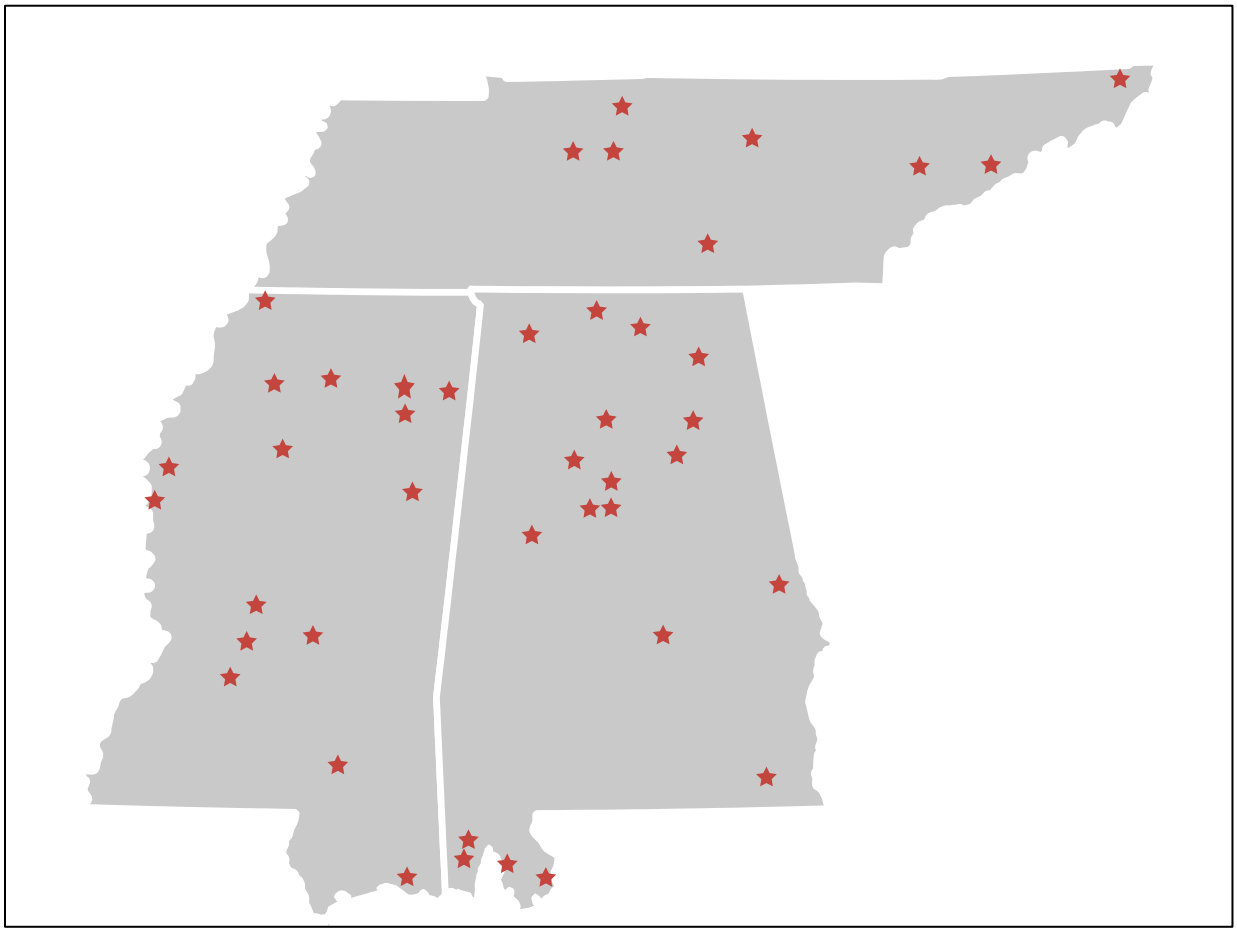
Study Setting

Main Study

The clinical partners consisted of 5 occupational therapists and 83 physical therapists and assistants at 43 clinics in three States: 19 in Alabama, 16 in Mississippi and 8 in Tennessee. Of these 43 clinic sites, 5 of the sites were considered small towns or rural, 10 were considered

micropolitan, and 28 were considered urban metropolitan. All clinics were considered outpatient rehabilitation clinics; however, 4 were located within hospitals, 38 were considered private practice and one site was a comprehensive MS center. **Figure 1** highlights the geographic location of clinical partners within each state (Alabama, Mississippi, and Tennessee).

Figure 1. Geographic Location of the TEAMS Clinical Partners (clinics= 43)



Enhancement Study

For the Enhancement study, participants received all testing through a teleassessment platform, followed by 20 synchronous telerehabilitation sessions in line with our original study design for the other two arms (TeleCAM, DirectCAM). A total of 8 physical therapists were involved in the testing and rDirectCAM delivery of the Enhancement study.

Participants

The study population included people with MS in the states of Alabama, Mississippi, and Tennessee. Interested participants were enrolled if they self-reported a diagnosis of MS, with a PDDS score between 0 and 7. The PDDS is an ordinal scale ranging from 0 (Normal) to 8 (Bedridden). It is a simple and practical tool for measuring self-reported disability status for people with MS[56-58] and has a strong correlation (*rs*=0.8) with the Expanded Disability Status Scale (EDSS)[56]. Additional inclusion criteria included ages of 18 to 70 years, able to use arms and legs for exercise, and physician clearance. Exclusion criteria included: 1) significantly impaired acuity that prevented seeing a tablet screen to follow home exercise program; 2) cardiovascular disease event within the past six months, severe pulmonary disease, renal failure; 3) active pressure ulcer; 4) currently pregnant; 5) within 30 days of receiving an exercise or rehabilitation program; 6) Godin Leisure Time Exercise Questionnaire (GLTEQ) score ≥ 24.

The GLTEQ contains two questions. The first question asks participants to report weekly frequencies of activities they perform at different intensities. A total weekly leisure activity is a sum of activity scores calculated by multiplying the weekly frequencies of strenuous, moderate, and light activities by 9, 5, and 3, respectively. The second question asks participants the

frequency of weekly leisure-time activities performed that are long enough to work up a sweat. The scale has been validated and used in previous research with people with MS[59,60].

Adverse events (AEs) were monitored following types defined by the Behavior Change Consortium of the National Institutes of Health[61], which included: falls, cardiovascular-related episodes, musculoskeletal-related events and health care use. The study was approved by the Institutional Review Board of the institution and was conducted in accordance with a clinical trial protocol registered at ClinicalTrials.gov.

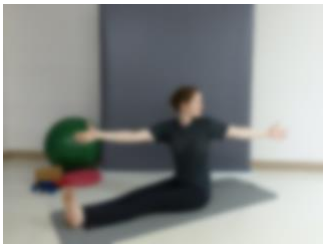
Interventions and Comparators or Controls

The two study arms received the same CAM intervention, and the comparison was between the on-site delivery (DirectCAM) versus the asynchronous delivery (TeleCAM). The CAM intervention is a 12 week intervention (3 month) composed of 20 sessions that were structured into two segments: the first 8 weeks consisted of two 1-hour sessions per week that included 3 components: 20-minute yoga exercise routine[62-69]; 20-minute Pilates exercise routine[70-74]; and a 20-minute routine of dual tasking and functional exercises that involved multifaceted training such as multidirectional upper extremity reaching while carrying on a conversation, and ball bouncing while tracking the ball with head and eye movements[75,76]. From weeks 9 to 12, the intervention transitioned to one session per week that focused on yoga and Pilates. The intervention was adapted and tailored toward the functional level of each participant, determined by the time it took for participants to complete the Timed 25-Foot Walk (T25FW) test. There were 4 functional levels provided. A participant who had a T25FW score of less than 6 seconds was assigned to TEAMS 1, a T25FW score of 6 to 7.99 seconds was

TEAMS 2, a T25FW score of more than 8 seconds was TEAMS 3, and if unable to complete the T25FW, TEAMS 4 was assigned. If the participant self-reported a diagnosis of osteoporosis and was able to complete the T25FW, they were assigned to TEAMS 3-OP. Those with a diagnosis of osteoporosis that were unable to complete the T25FW were assigned to TEAMS 4-OP. TEAMS 1 and TEAMS 2 completed the intervention exercises in prone, supine, long sitting, standing, and quadriped. TEAMS 3 and TEAMS 3-OP completed the exercises in a chair and with an assistive device for support in standing. TEAMS 4 and TEAMS 4-OP completed all the exercises in a chair. All TEAMS levels included ways to tailor the exercises to better align each participant's functional level. **Figure 2** demonstrates examples of the exercises and adaptation of the CAM intervention.

Figure 2. Examples of the CAM Intervention Exercise and Adaptation

Program Component	Sample Content	Modification Examples
Yoga Exercise (20 minutes)	**Tree Pose** *Standing:* 1) Begin in standing. Bring shift your weight to your left side and bend your right knee sliding your foot against your inner thigh to your farthest range 2) Bring your hands into prayer position with your thumbs resting on your chest. 3) Hold this pose for four breaths.  *Chair:* 1) Begin sitting in a chair. Extend your left leg out straight and shift your weight to your left side. 2) Bend your right knee and slide your foot against your inner thigh to your farthest range. 3) Bring your hands into prayer position with your thumbs resting on your chest. 4) Hold this pose for 2 breaths	<u>Lower difficulty:</u> Bring the right foot to the left ankle to rest on the toes without sliding up. <u>Higher difficulty:</u> Extend arms out above instead of in prayer position.

Pilates Exercise (20 minutes)	**Spine Twist** *Long sitting:* 1) Start with sitting on the floor with legs extended in front of you and arms open to your sides at shoulder height. 2) Exhale and rotate your torso to the right side (twist); make sure your head and arms follow 3) Inhale and return to starting position. 3) Exhale and rotate your torso to the left side (twist); make sure your head and arms follow *Chair:* 1) Sitting in a chair with both feet on the floor and arms open at your sides at shoulder height. 2) Exhale and rotate your torso to the right side (twist); make sure your head and eyes follow as well. 3) Inhale and return to starting position. 4) Exhale and rotate your torso to the left side (twist); make sure your head and eyes follow as well. 	Lower difficulty: Cross arms across chest instead of extended at sides. Higher difficulty: 1) Utilize a TheraBand with arms extended for long sitting exercise challenge 2) Sit on the Pilates disc for chair exercise challenge. 
Dual Tasking Exercise (20 minutes)	**Dual Tasking:** *Vision:* Smooth pursuits and Vestibular Ocular Reflex exercises; *Cognition:* Process Speed exercises through alphabet object naming and simple mathematic tasks; *Gross and fine motor coordination:* Dysdiadochokinesia exercises 	Lower Difficulty: 1) Perform same tasks while sitting supported in chair. Higher difficulty: 1) Perform same tasks while sitting on physioball or Pilates Disc; 2) Perform same tasks while standing in partial tandem, tandem or on a Pilates Disc.

Participants were given exercise equipment after baseline testing for the Main Study. Both arms received the same equipment based on their TEAMS level. The TeleCAM arm received a tablet and tablet stand in addition to the exercise equipment. The DirectCAM received the exercise equipment to use at home, and the clinic kept a set of equipment for use at onsite visits. **Table 2** provides an overview of the CAM intervention for the Main study.

Table 2. Exercise Equipment by TEAMS Level

Equipment	TEAMS Level 1	TEAMS Level 2	TEAMS Level 3	TEAMS Level 4
Racquet Ball	√	√	√	√
Yoga Strap	√	√	√	√
Theraband	√	√	√	√
Pilates Disc	√	√	√	√
Sliders	√	√	√	√
Yoga Mat	√	√	√	√
Yoga Block	√	√	√	
Physioball	√	√	√	
Half Roll	√	√		
Tablet (TeleCAM)	√	√	√	√
Tablet Case (TeleCAM)	√	√	√	√
Tablet Stand (TeleCAM)	√	√	√	√

TEAMS: Tele-Exercise and Multiple Sclerosis, CAM: Complementary and Alternative Medicine, T25FW: Time 25 Foot Walk score. TEAMS level 1: T25FW <6 seconds, TEAMS level 2: T25FW 6-7.99 seconds, TEAMS level 3: T25FW >8 seconds, TEAMS level 4: unable to complete T25FW.

For the Enhancement study, participants received an additional set of equipment (in addition to the exercise equipment) that was shipped to their home for completing the teleassessment protocol. The teleassessment equipment provided to participants in the Enhancement study is outlined in Table 3. For safety considerations, rDirectCAM participants who had a PDDS score of 5, 6, and 7 were not given any tests related to ambulation. Hence, two different sets of equipment were provided: one for participants with a PDDS score of 0-4 and another for participants with a PDDS score of 5-7.

DirectCAM. DirectCAM participants were asked to work with a therapist at the assigned clinic to complete the 20 one-hour training sessions as described in **Figure 2**. A therapists training manual and a DirectCAM session checklist were used to ensure intervention fidelity across therapists (see Appendices A and B). At the end of the 12-

week intervention, the DirectCAM participants received a set of instructions and handouts (see Appendix C) with photos demonstrating the exercises and were asked to continue the program in their home for up to one-year post-intervention initiation (i.e., 9 months after the completion of the intervention delivered at the clinic).

TeleCAM. The TeleCAM participants received the same intervention content as DirectCAM. The only difference was that the TeleCAM participants received the intervention via exercise videos preloaded to a tablet that was given to them upon enrolling in the study. The exercise videos were filmed with a person with MS instructing and demonstrating the CAM program. Each TeleCAM participant began with one onsite training session with a study therapist. In the training session, participants were shown how to turn the tablet on and off to watch the videos, perform the basic exercises correctly, and use the yoga and Pilates equipment. Participants were asked to bring a family member or friend with them in case they needed assistance using the tablet. The TeleCAM participants were contacted by an Interactive Voice Response (IVR) system-initiated call at their preferred dates and times (aligned with the general intervention timeline) to promote adherence to the yoga/Pilates exercise and neurorehabilitation practice. The IVR system is an automated communication system that includes computers, specialized phone hardware, and operating software to allow immediate computer-assisted interaction with participants by telephone. It was developed during our research team's past physical activity research and in the current study was adapted to address the specific concerns of individuals with MS (physical activity progress, pain, fatigue, and key motivational variables from the SCT). The IVR calls were sent to only

TeleCAM participants weekly for the first six months, and then biweekly for months 6-12 to monitor physical activity, pain, fatigue, and quality of life for SCT variables.

rDirectCAM. The rDirectCAM was developed in response to the Covid-19 pandemic and was the Enhancement study intervention. Participants enrolled in rDirectCAM received the same intervention as DirectCAM. The only difference was that the rDirectCAM participants utilized a remote platform at home to work synchronously with a therapist.

Study Outcomes

For the Main Study, all data collection for both DirectCAM and TeleCAM participants was conducted using standardized protocols and was performed by a therapist or therapist assistant at each clinic site who was not involved in the intervention delivery. In addition, demographic (age, sex, and race) and clinical characteristics (health history, disease severity via PDDS, heart rate, and blood pressure)[77] were assessed.

Primary Outcomes. The primary outcomes of interest to our stakeholders included changes in fatigue, pain, quality of life, and physical activity. Fatigue was assessed using the Modified Fatigue Impact Scale, pain assessed using SF-36 pain subscale and quality of life via SF-36 questionnaire's physical component and mental component scores. Physical activity was measured using the GLTEQ. Note that the scales and questionnaires were self-reported by the participants.

Modified Fatigue Impact Scale (MFIS). The MFIS consists of 21 items that assess the effects of fatigue related to physical, cognitive, and psychosocial functioning. It is a

modified version of the Fatigue Impact Scale. The scale has been widely used in MS research[78].

SF-36. The SF-36[79] is a widely used questionnaire that assesses aspects of health-related quality of life. It contains eight subscales including: energy/fatigue, physical functioning, bodily pain, general health perceptions, role limitations due to physical health problems, role limitations due to personal or emotional problems, social functioning, and emotional well-being. Scores range from 0 to 100, with lower scores indicating more disability. From these eight scales, 2 composite scores: physical component score (PCS) and mental component score (MCS) were calculated by multiplying each subscale with specific weights.[80] These 2 scores are measuring QOL.

Secondary Outcomes. The secondary outcomes were changes in functional scores for balance, endurance, gait, and strength. These tests were assessed by therapists at the clinic for both DirectCAM and TeleCAM participants.

Berg Balance Scale (BBS). The BBS is administered by a single rater who scores an individual's balance performance from a value of 0 (cannot perform) to 4 (normal performance) on a total of 14 tasks [81]. Tasks incorporate several activities related to static balance such as sitting to standing, retrieving an object from the floor, standing with eyes closed, and standing on one foot. The test was administered with a ruler, two chairs (height: 45–47 cm; one with armrests and one without), a footstool, a 15-foot walkway, and a stopwatch. Previous research has reported a minimum detectable change of seven points in multiple sclerosis[82].

Five Time Sit To Stand (FTSTS). The FTSTS was used to assess balance, it measures the amount of time it takes for an patient to transfer from a seated position to standing position and back to sitting five times. The FTSTS test has been shown to effectively assess lower extremity strength in individuals with MS [83]. Furthermore, its reliability and validity have been established in various other neurological disorders such as stroke and Parkinson's disease[84].

Six Minute Walk Test (6MWT). The 6MWT was used to assess walking endurance, in which total distance traveled over 6-minute period was recorded in meters[85,86]. The test was administered in a hallway with an unobstructed 50-foot path marked with two cones. Perception of fatigue was measured and recorded with the BORG Rating of Perceived Exertion (RPE) before and after the test. The test has been reported to have excellent test-retest (ICC = 0.97) in the stroke population and people with MS[85,87] and has known validity for use in measuring endurance in persons with MS[85].

Timed Up and Go (TUG). The TUG test was used to assess mobility and balance, which involves recording the time required to rise from a chair, walk to a 3-meter mark, turn around, walk back to the chair, and sit down[88,89]. The test is a valid test of mobility, balance, and walking ability in people with MS[90], with excellent test-retest reliability (ICC = 0.96)[91].

Grip Strength Hand Dynamometer. Hand grip strength was measured using a hand dynamometer with the dominant and non-dominant hands. The mean of three trials for each hand was used as the actual score. The text was conducted with the participant siting in a chair, shoulder adducted and neutrally rotated, elbow by their side and flexed

to a 90° angle, and forearm and wrist in neutral position. Grip strength has previously shown a high interrater reliability and test-retest reliability in persons with MS[92].

Time 25 Foot Walk. T25FW is commonly administered within the MS functional composite (MSFC), and it has also been employed independently in clinical investigations[93,94]. It is a quantitative mobility and leg function performance test based on a timed 25-walk. The patient is guided to a designated starting point on a clearly outlined 25-foot walk and is advised to walk the distance as swiftly as they can while ensuring safety. Timing begins when the instruction to commence is given and concludes once the individual reaches the 25-foot mark. The process is repeated immediately with the individual walking back the same distance. Participants are permitted to utilize assistive devices during this activity.

Mediators. Four mediators were assessed to address the tertiary aim.

Social provision scale (SPS). The SPS was used to assess the degree to which a participant perceived his or her social relationships and various dimensions of social support[95]. The scale has been validated in people with MS[96].

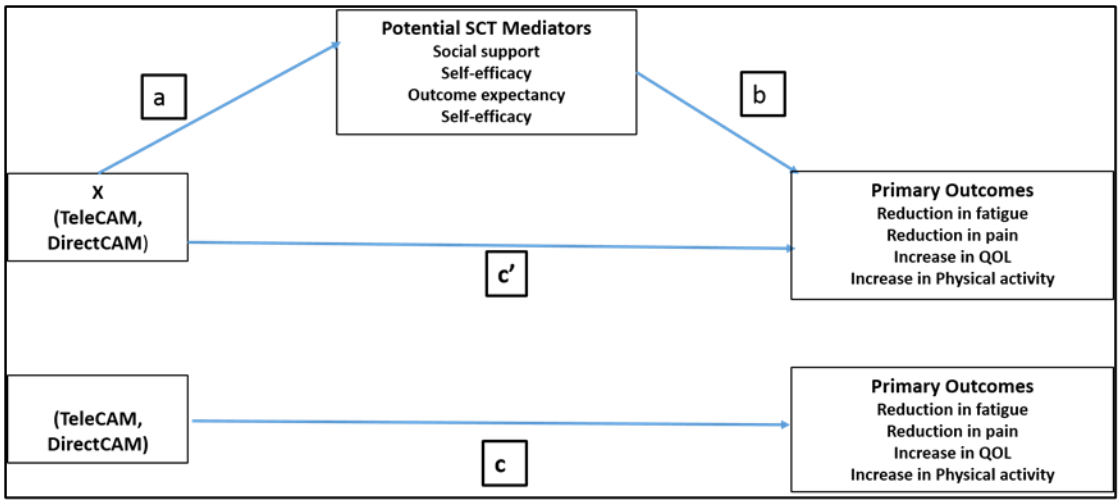
Exercise Self-Efficacy (EXSE): The EXSE measures self-efficacy related to physical activity and consists of 6 items that assess an individual's belief in their capability to participate in moderate physical activity for at least 20 minutes, three times a week, within one-month intervals. Participants rated each item on a scale from 0 (not confident at all) to 100 (completely confident) and the scores were averaged to create a composite score ranging from 0 to 100. Higher scores indicate higher levels of confidence in regularly engaging in physical activity[97].

Exercise Goal-Setting Questionnaire (EGSQ): The EGSQ comprises 10 items that assess goal-setting behaviors for exercise. Participants rated each item on a 5-point scale ranging from 1 (does not describe) to 5 (describes completely). The scores of the individual items were then added together to calculate an overall score, ranging from 10 to 50. Higher scores on the EGSQ indicate a greater inclination towards setting goals for engaging in physical activity[98].

Multidimensional Outcome Expectations for Exercise Scale (MOEES). The MOEES was used to assess outcome expectations for exercise. The scale consists of 15 items representing three subdomains of outcome expectations: physical, social, and self-evaluative. Within these subdomains, six items pertain to physical outcomes, four items to social outcomes, and five items to self-evaluative outcomes. Participants rated each of the 15 items on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The scores for each subscale were then added together, resulting in three separate scores reflecting physical, social, or self-evaluative outcome expectations. Higher scores in each subscale indicate a stronger belief in the expected positive outcomes related to physical, social, or self-evaluative aspects of exercise[99].

Figure 3 shows mediation model with the effect of 4 Social Cognitive Theory (SCT) construct variables as mediating variables.

Figure 3. Illustration of the mediation analysis:



X: the independent variable, c: the coefficient relating the independent variable and the dependent variable, c': the coefficient relating the independent variable to the dependent variable adjusted for the mediator, b: the coefficient relating the mediator to the dependent variable adjusted for the independent variable, a: the coefficient relating the independent variable to the mediator.

Table 3 provides an overview of the primary, secondary, and exploratory outcomes and the frequency and mode of data collection. To further tailor our intervention, the T25FW test was administered. In addition, the Dynamic Gait Index (DGI) was administered at clinic sites that had uniform steps to assess the test and compare its utility with BBS[100]. However, these were exploratory outcomes and were not part of our original proposal; Hence, they were not accounted for while developing power calculations for the study in terms of the number of multiple comparisons for the family wise error rate.

Table 3. TEAMS Study Outcome Measures

Variables	Instrument	Role, Frequency, Mode of Collection
Fatigue	Modified Fatigue Impact Scale SF-36 PCS SF-36 MCS	Primary Outcome Baseline, 3rd, 6th and 12th month Self-reported
Pain		
Quality of life		
Physical activity	Godin Leisure Time Exercice Questionnaire[101]	
Balance	Berg Balance Scale[102]	Secondary Outcome

	Five Time Sit to Stand[83]	Baseline, 3rd, 6th and 12th month At the clinic (for Main study)
Gait	Timed Up & Go (TUG)[103-105] Time 25 Foot Walk[93,94]	
Endurance	6-minute Walk Test (6MWT)[105-111]	
Strength	Hand-held Dynamometer/Grip Strength[112-114]	
Quality of life	SF-36 Physical functioning, Pain, Energy/Fatigue, Emotional wellbeing, Role limitation due to physical health, Role limitation due to emotional health, Social functioning, General health	Exploratory outcomes Baseline, 3rd, 6th and 12th month Self-reported
Gait	Dynamic Gait Index Scale	Exploratory outcomes Baseline, 3rd, 6th and 12th month Objective outcomes
Outcome expectations for exercise	Multidimensional Outcome Expectations for Exercise Scale (MOEES)[115]	Mediators Baseline, 3rd, 6th and 12th month
Exercise self-efficacy	Exercise Self-efficacy Scale[116,117]	
Social support for exercise	Social Support and Exercise Survey[118]	
Exercise self-regulation	Exercise Goal-setting Scale[119]	

IVR: Interactive Voice Response.

Enhancement Study

Outcome measures for the Enhancement study were collected using teleassessment via the remote platform. To ensure strong fidelity across therapists, each therapist underwent extensive training on use of the technology, risk management for remote assessment and intervention, and teleassessment and intervention protocols. The goal of the protocols was to mirror onsite clinic assessments and intervention. The measures utilized were based on the

participants' PDDS score. Participants with a PDDS score between 0-4 (indicative of minimal mobility disability) completed the following measures: Grip strength test, Five Times Sit to Stand test, TUG, and BBS. Participants with a PDDS score between 5-7 completed all measures except the TUG and BBS due to safety concerns (potential risk of falling) of performing these tests at home. The teleassessment equipment provided to participants in the Enhnacement study is outlined in **Table 4**. For safety considerations, rDirectCAM participants who had a PDDS score of 5, 6, and 7 were not given any tests related to ambulation. Hence, two different sets of equipment were provided: one for participants with a PDDS score of 0-4 and another for participants with a PDDS score of 5-7. More details of the teleassessment are presented in the Appendix D.

Table 4. Teleassessment Equipment for the Enhancement Study

PDDS Level	Equipment
PDDS 0-4	Chromebook CAMRY Digital Hand Dynamometer Grip / 90 Kgs A&D Medical Wrist Blood Pressure Monitor SumVibe Soft Tape Measure 120 Inches American Challenge Soccer Mini Disc Cone Wall tape Paper indicating 5" line
PDDS 5-7	Chromebook CAMRY Digital Hand Dynamometer Grip / 90 Kgs A&D Medical Wrist Blood Pressure Monitor

PDDS: Patient Determined Disease Steps.

Potential covariates

In addition to age, sex, education-level, height, and race, which were collected at baseline, information on all participants' time-varying covariates were collected through a questionnaire at baseline and the three follow-up time points. These covariates included: weight, type of MS, prescribed and non-prescribed medications, distance from the clinic

(address), caregiver support, assisted devices, PDDS. The type of MS, PDDS, age and geographic locations based on participants zip code (urban/non-urban) were our prespecified moderators for the study outcomes.

Sample Size Calculations and Study Power

For the Main study an *a priori* power analysis was conducted based on sample size estimation for longitudinal designs with attrition with at least 30 sites. Based on the conservative power and sample size calculations, the project was powered for a minimum of 533 participants completing the study with an assumption of a 35% attrition rate; therefore, a total estimated sample size of 820 was needed to detect a between group effect. We assumed intent to treat analyses, 80% power, two-sided family-wise error rate of 0.05, intra-class correlation of 0.05, and a correlation of 0.3 among repeated measures. We also assumed an attrition rate of 35% and a small minimally detectable effect size of 0.2. We make a conservative assumption that the correlation between repeated measures would be 0.3. We have four primary and four secondary outcomes and hence the sample size was estimated after accounting for multiple testing and with a type 1 error rate of 0.00625 for each hypothesis test.

For the Enhancement study, we assumed ANCOVA with a pre-post correlation of 0.7, 80% power, one-sided test with statistical significance of 0.05, and attrition rate of 25%. Further, these sample size calculations assumed a non-inferiority margin of -4 units for MFIS, where fatigue is one our key primary outcome identified by the stakeholders in our project. A change of 4 or greater has been found clinically meaningful for quality of life. Based on these assumptions a sample size of 55 completers in rDirectCAM were needed and with an attrition

rate of 25%, we needed 74 participants in rDirectCAM.

Time Frame for the Study

For the Main study, TEAMS launched its participant recruitment in October 2017 and ended participant enrollment in 2019. **Figure 4** shows the timeline for a few of the major milestones completed during the study period. Recruitment for the rDirectCAM group (Enhancement study) was not necessary as we had interested participants who already signed up the study via our website or through our clinic partners but were not able to enroll before the Main study needing to be ceased due to the pandemic. We established a protocol for conducting the teleassessments and tele-training and developed a structured set of procedures for equipment purchase and shipping/scheduling based on the Enhancement supplement received in May 2020. We began participant screening for the Enhancement study on September 10, 2020, with 100% enrollment completed on March 22, 2021. The study was closed on November 1, 2021.

Figure 4. Major Milestone Completion Timeline of the Main Study

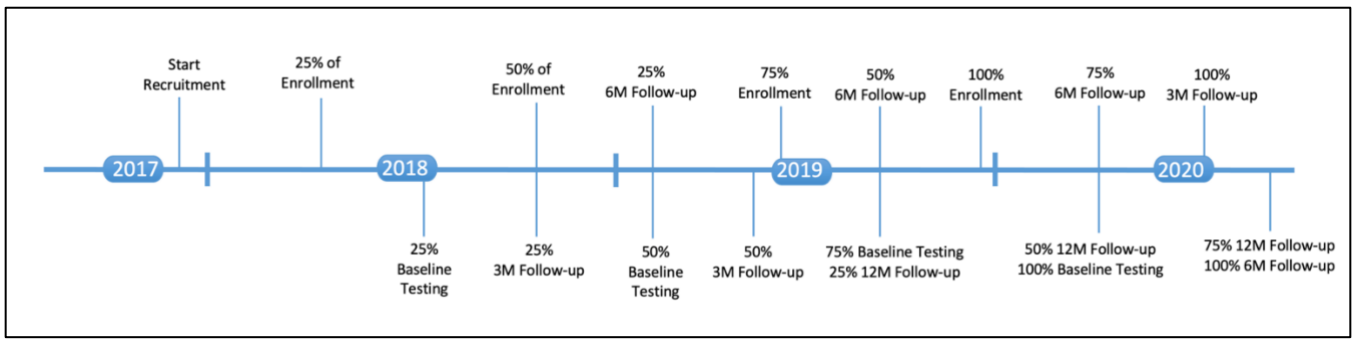


Table 5 provides an overview of the timeline for the data collection, intervention, and regular check-ins of the Main and Enhancement studies.

Table 5 Overview of the Data Collection, Intervention, and Check-ins Timeline for the Main and Enhancement Studies

	Main Study	Main Study	Enhancement
Assessment time point	**DirectCAM**	**TeleCAM**	**rDirectCAM**
Month 0	Baseline assessment (on-site)	Baseline assessment + Training (on-site)	Baseline assessment (real-time via video conference)
Months 1 to 2	Biweekly sessions (on-site)	Biweekly new videos (at-home) + IVR calls weekly	Biweekly sessions (real-time via video conference)
Month 3	Weekly sessions (on-site)	Weekly new videos (at-home) + IVR calls weekly	Weekly sessions (real-time via video conference)
End of Month 3 (post intervention: 12 weeks)	3 month assessment (on-site)	3 month assessment (on-site)	3 month assessment (real-time via video conference)
End of Month 6	6 month assessment (on-site)	6 month assessment (on-site)	6 month assessment (real-time via video conference)
End of Month 12	1 year assessment (on-site)	1 year assessment (on-site)	NA
Months 4 to 12	Continue to exercise (at-home)	Continue to exercise (at-home) + IVR calls 4-5 months weekly/ 6-12 months biweekly	Continue to exercise (at-home)

Data Collection and Sources

Study data were collected by the therapists at the clinic sites via paper and scanned into a secure cloud storage environment Box approved for UAB investigators. The data were then transferred and managed using REDCap (Research Electronic Data Capture). REDCap is a secure, web application designed to support data capture for research studies, providing user-friendly, web-based case report forms, real-time data entry validation (e.g., for data types and range checks), audit trails and a de-identified data export mechanism to common statistical packages (SAS version 9.4[120], R version 4.3.2). Network transmissions (data entry, survey submission,

web browsing, etc.) in REDCap are protected via Secure Sockets Layer (SSL) encryption. All data was quality checked by different team personnel to detect and correct data entry errors.

Missingness in the weekly feedback, monthly motivational, and clinic assessment data stored in the REDCap database was tracked periodically. Data were imported into the REDCap server on a weekly basis and queried for missingness. Reminder calls were sent to participants to reduce missed appointments and to remind participants to respond to IVR calls in TeleCAM. If more than 2 data points were missed, the study team called both therapists and participants to inquire why, and to reengage as many as possible and retrieve the missing data. Clinics were provided with checklists to ensure that all measures were collected at each assessment session in a consistent order according to study protocol. Checklists minimized data loss due to technical difficulties. Missed appointments were rescheduled immediately**.** Reasons for drop-out (and whether drop-out was therapist or participant-initiated) were documented in detail in Box, REDCap and in the monthly CONSORT diagram**.** To minimize loss to follow-up, multiple types of contact information were collected. Protocols were developed by our stakeholders for contacting participants in an approachable nature throughout the study through both telephone and emails.

Analytical and Statistical Approaches

Main Study

Our analysis was following ITT protocol as all participants assigned to the randomized clinics and completed baseline assessment were included in the analysis.

Our primary hypotheses were about comparing the change in the 11 outcomes (listed in **Table 4**) between baseline and 12th month follow-up across the two arms. We considered these 11 hypothesis tests for the 11 outcome variables as part of a single family of hypotheses. We employed the Holm-Bonferroni method[121] for controlling the overall family wise error rate and computed adjusted *p*-values. For each outcome we used a linear mixed effects model, with TeleCAM arm as the reference category, and accounting for the correlated nature of the data due to the nesting of participants within a clinic as random effects. Crude models included fixed effects included both Arms (TeleCAM, DirectCAM), Time (baseline [reference], 3 month, 6 month, 12 month), an Arms X Time interaction, and accounted for randomization strata (clinic location being urban vs. non-urban). The interaction of Arms and Time variables was the primary effect of interest since it quantified the differences within Arm change in the outcome from baseline over time of the two Arms. For example, based on the fitted model that included the interaction between Arm and Time, the difference in change over time from baseline to 12 months between DirectCAM and TeleCAM were estimated. Contrast statements were used to test the null hypotheses in conjunction with the fitted model coefficients and standardized differences in the average treatment effect reported. The estimates of the change in health and function outcomes for each arm and their 95% confidence interval were also reported. As a sensitivity analysis, similar models additionally adjusted for sex, race, and post-COVID data collection measure indicated were also ran. Note that, since post-COVID, data collection for patient reported outcomes was conducted remotely for remaining participants, the former was also included as a covariate in the adjusted model. Next, missingness in primary and secondary outcomes were addressed using multiple imputation technique (MCMC) via PROC MI in SAS,

assuming missing at random. Additional sensitivity analyses were conducted using wild bootstrap to assess the robustness to model assumptions.

The estimated changes between groups and within groups were then compared to minimally clinically important difference (MCID) published in literature for the studied outcomes among people with MS. For primary outcomes, the published MCID was for MFIS= 4[122] and SF-36 pain score = 4.9[123]. For secondary outcomes, MCID for 6MWT= 21.56[124], BBS= 3[125], TUG= 10.6, FTSTS= 13[126] and grip strength= 10.5[127]. For exploratory outcomes, MCID was published for SF-36 physical functioning (MCID= 7.1[123]).

Aim 2 was about evaluating pre-specified confirmatory HTE analyses by testing the role of 4 prespecified moderators (age, geographic location/residence based on participants zip codes, type of MS and severity of MS) on each of the primary self-reported outcomes: MFIS, SF-36 MCS, SF-36 PCS, SF-36 pain domain scale, and GLTEQ. We had 4 families of hypothesis tests within this aim – one for every moderator with 2 levels (PDDS 0-3 [without assistive devices] and PDDS 4-7 [with assistive devices]). Our model was like the one used in the primary aim except that we augmented the model with an additional two-way interaction term between moderator (e.g., education level) and Arm (moderator X Arm), and a 3-way interaction term between moderator x Arm x Time. When a 3-way interaction was observed, post-hoc tests were performed to determine the differential influence of the moderator to one treatment versus another, controlling for family-wise error utilizing the Holm-Bonferroni method. We expected no greater than 11 post-hoc tests per pre-specified moderator and used contrast statements to achieve this. Our missing data strategy was similar to that in the primary aim.

We examined the role of SCT construct variables (social support, outcome expectancies, self-efficacy, self-regulation) as mediating variables between the intervention and primary outcomes. As part of our tertiary aims, we had 4 families of hypotheses, one for each mediating variable listed in **Table 4**. Each family of hypotheses had 5 hypotheses tests, one for each primary outcome. We applied the product of coefficients test framework as our primary approach to compute the mediated effects for each of the primary outcomes as suggested by MacKinnon[128,129]. Additionally, we did mediation analysis for the 6 secondary and the 7 exploratory outcomes. Note that these analyses were conducted using multiple imputed datasets created for aim 1 sensitivity analyses.

Enhancement Study

We used a quasi-experimental design to compare rDirectCAM with DirectCAM. Our hypothesis posits that the observed difference in outcome effects between rDirectCAM and DirectCAM is less than a predetermined non-inferiority margin (-Δ). These margins were selected based on the MCID or minimal detectable change (MCD) of the studied outcomes among MS population published in literature (MFIS= 4[122], BBS= 3[125], TUG= 10.6[130], FTSTS= 13[126], Grip-strength= 10.5[127], SF-36 Physical functioning= 7.1[123] and SF-36 pain= 4.9[123]).

Specifically, we hypothesized that the effect of rDirectCAM was not inferior to that of DirectCAM by a margin that was considered clinically acceptable, thereby affirming the non-inferiority of rDirectCAM in comparison to DirectCAM. In other words, if the resulting *p*-valuewas significant, this would mean that we rejected the null hypothesis which stated that the difference in the effect between two arms were more than the non-inferiority margin. Statistical significance level of 0.05 and at confidence interval 95%, same as superiority trials,

was applied to determine non-inferiority.[131] To achieve a balance between the study arms, several analyses involved regression-based parametric adjustment measures that are sensitive to model misspecification. Propensity scores are another widely used measure to achieve balance or matching between the study arms. These scores model the probability of being assigned to the given treatment conditional on observed baseline covariates obtained via measures including matching, inverse probability treatment weighting (IPTW), and stratification and covariate adjustment. Matching-based measures such as nearest neighbor matching algorithm often led to study participants being excluded when a match was not available. This may have potentially biased the effect estimates. When this occurred, we utilized stabilized IPTW, a non-parametric balancing strategy, to achieve balance between the two study arms (rDirectCAM and DirectCAM). Using this weighted analysis, we modeled the causal parameter as the inverse of the conditional probability of receiving the treatment given the observed covariates, a technique that provides asymptotically unbiased estimates and uses data on all available study participants across the two arms.

Our general analytic approach was to leverage the linear mixed modeling framework to account for repeated nature of data. Appropriate contrasts were applied using the fitted model. Sensitivity analyses were performed to assess distributional and other modeling assumptions. Missing data was imputed using multiple imputation where necessary. The non-inferiority ($-\Delta$) of the rDirectCAM was assessed by comparing the lower limit of confidence interval for each outcome to its corresponding margin so that if the lower bound was above this margin, then the rDirectCAM was deemed non-inferior [132]. For the MFIS and TUG assessments, wherein an elevated score signified a deterioration in the outcome, we juxtaposed the upper limit of the

confidence interval against a predefined margin. Non-inferiority was deemed established when the upper bound was less than $-\Delta$[133].

Changes to the Original Study Protocol

Research Plan Changes:

1. January 12, 2018

   a. Increased the number of Drayer Physical Therapy clinic sites from 24 to 25 due to Drayer's purchase of the Fulton, MS clinic.

2. May 16, 2019

   a. Four Upstream Rehabilitation Inc. sites (Wiggins, Smyrna, Prattville and Saltillo) were replaced with four sites due to contamination issues, site not opening, and low recruitment opportunities. The replaced sites included two new sites (Forest, MS and Bristol, TN) under Upstream Rehabilitation Inc. and two sites (Phenix City, AL and Grenada, MS) under Encore Physical Therapy.

   b. Five new sites were added, with two as study sites and three as satellite sites. The two newly added study sites were Ruleville, MS under North Sunflower Medical Center and Greenville, MS under River City Rehabilitation. These two sites were added to serve the Delta area of Mississippi. The three newly added satellite sites were Ashland City, TN and Cookeville, TN under Upstream Rehabilitation Inc. and Mobile, AL under Encore Physical Therapy. Satellite sites were added to help with recruitment and testing needs. Therefore, the number of participating clinic sites increased to 43, with three of the sites serving as satellite sites.

3. March 24, 2020

a. Our stakeholder panel and the research team determined that *meeting physical activity guidelines* should be added to the inclusion/exclusion criteria to ensure that the intervention being delivered was the catalyst for change and not another exercise routine. Therefore, the *Godin Leisure Time Exercise Questionnaire (GLTEQ)* was added to the exclusion criteria and a GLTEQ score ≥ *24 was used as an indicator of individuals meeting physical activity guidelines* during the screening process. This addition was approved by our IRB on July 25, 2017.
b. The EDSS (1-7) was changed to PDDS (0-7) as an inclusion criterion. The rationale for this change was discussed with our stakeholder panel and research team upon realization that very few people with MS knew their EDSS score. This change was approved by our IRB on July 25, 2017.
c. To honor our commitment to participants enrolled in the study and provide them with the intervention and follow-up assessments remotely during Covid-19, we requested approval for the following modifications:
    i. Contacted all therapists and participants to suspend onsite visits and requested their participation in completing the intervention (rDirectCAM) and/or performing all testing remotely (teleassessment).
    ii. The following steps were followed for offering rDirectCAM to participants:
        - For those enrolled in DirectCAM who had started the Intervention (n-=25):

    - Sent participants a tablet with access to a Zoom link for rDirectCAM.
    - Trained therapists to deliver rDirectCAM remotely through the Zoom platform.
- For participants who did not have internet access, we offered a phone call option in which the therapist went through the exercises with the participants via phone, utilized the provided DirectCAM booklet (e.g., “Ok. Now turn to page 3 and let’s do the Warrior pose.”)

4. August 20, 2022
    a. The Covid-19 Enhancement study was added into the research plan.

Research Analysis Changes:

The seven SF-36 subscales were studied as exploratory outcomes. This was done because they captured all aspects of QOL, and we were interested to explore what was the effect of intervention on each domain.